Constraints, Covariances and χ^2 in Solve

John M. Gipson

March 5, 1997

Abstract

In this note I look at the effect of constraints on and χ^2 in solve. We are currently not using the correct formula. This leads to an error in χ^2 at the 1-10% level.

The purpose of this memo is to derive various covariance formulas. Upper case italic letters will denote matrices or vectors. Assume that we have some model which depends linearly on *nparam* parameters A, and that we have *nobs* observations Y. The A's and the Y's are related by:

$$Y = FA \tag{1}$$

where F is a known matrix of dimension $nobs \times nparam$ (the partials in solve, or the design matrix). In the model is correct, and there is no noise, this equation is exact. In the presence of noise this equation is modified to:

$$Y = FA + \varepsilon \tag{2}$$

Here ε is the noise associated with a given observation. As stated this equation may have one, many or no solutions. Following standard practise, instead of solving (2) directly, we minimize the weighted residuals:

$$\left(Y^T - A^T F^T\right) W \left(Y - FA\right) \tag{3}$$

with respect to A. Here

$$\langle \varepsilon_l \varepsilon_m \rangle = W^{-1}_{lm} \tag{4}$$

i.e., W is the covariance of the observations. The condition that 3 be a minimum is easily found to be:

$$F^T W F A = F^T W Y \tag{5}$$

or alternately:

$$N_0 A = B_0 \tag{6}$$

which implicitly defines

$$N_0 \equiv F^T W F \tag{7}$$

$$B_0 \equiv F^T W Y \tag{8}$$

In general equation (5) is not invertible because of degeneracies in the normal equations. In solve this is circumvented by adding a constraint matrix to the left hand side of (5). The constraint equation can also be added to reflect a priori knowledege, or to make the solution "better behaved". In any case, the normal equations are modified as follows:

$$\left(F^T W F + G\right) A = F^T W Y + Z \tag{9}$$

or

$$N A \equiv B \tag{10}$$

where

$$\begin{aligned} N &= N_0 + G & (11) \\ B &= B_0 + Z & (12) \end{aligned}$$

which has solution:

$$\begin{aligned} A &= \left[F^T W F + G\right]^{-1} F^T W Y & (13) \\ &= [N_0 + G]^{-1} B & (14) \end{aligned}$$

The residual vector is defined as:

$$\begin{aligned} \Delta &\equiv Y - F A & (15) \\ &= Y - F N^{-1} F^T W Y - F N^{-1} Z & (16) \\ &= (I - F N^{-1} F^T W) Y - F N^{-1} Z & (17) \end{aligned}$$

this measures the degree of mis-closure. If the data were noiseless, and the model perfect, and there were no constraints, this vector would vanish. We are interested in the covariance of this vecotr. First we need to calculate the dependence of the residual vector on the observations. We find:

$$\frac{\partial \Delta_j}{\partial Y_k} = \left[I - F N^{-1} F^T W\right]_{jk} \tag{18}$$

Now, the covariance of this vector is given by:

$$\begin{aligned} [C_\Delta]_{j,k} &= \sum_{l,m} \left\langle \frac{\partial \Delta_j}{\partial Y_l} \varepsilon_l \frac{\partial \Delta_k}{\partial Y_m} \varepsilon_m \right\rangle & (19) \\ &= \sum_{l,m} \frac{\partial \Delta_j}{\partial Y_l} \frac{\partial \Delta_k}{\partial Y_m} \langle \varepsilon_l \varepsilon_m \rangle & (20) \end{aligned}$$

Using (4) we find

$$\begin{aligned} C_\Delta &= \left[I - F N^{-1} F^T W\right] W^{-1} \left[I - W F N^{-1} F^T\right] & (21) \\ &= W^{-1} - 2 F N^{-1} F^T + F N^{-1} N_0 N^{-1} F^T & (22) \\ &= W^{-1} - F N^{-1} F^T - F N^{-1} C N F^T & \end{aligned}$$

In the absence of constraints the last term vanishes, and we have:

$$\begin{aligned} C_\Delta &= W^{-1} - F N_0 F^T & (23) \\ &= W^{-1} - F \left[F^T W F\right]^{-1} F^T & (24) \end{aligned}$$

Consider

$$W^{\frac{1}{2}}C_\Delta W^{\frac{1}{2}} = I_{nobs} - W^{\frac{1}{2}}FN^{-1}F^TW^{\frac{1}{2}} - W^{\frac{1}{2}}FN^{-1}CN^{-1}F^TW^{\frac{1}{2}} \tag{25}$$

where I_{nobs} is the $nobs \times nobs$ unit vector. The diagonal elements on the LHS are just

$$\left[W^{\frac{1}{2}}C_\Delta W^{\frac{1}{2}}\right]_{jk} = \frac{\Delta_j^2}{\sigma_j^2} \tag{26}$$

hence

$$\begin{aligned}
\chi^2 &= \sum_j \frac{\Delta_j^2}{\sigma_j^2} & (27)\\
&= \mathrm{Tr}W^{\frac{1}{2}}C_\Delta W^{\frac{1}{2}} & (28)\\
&= \mathrm{Tr}I_{nobs} - \mathrm{Tr}W^{\frac{1}{2}}FN^{-1}F^TW^{\frac{1}{2}} - \mathrm{Tr}W^{\frac{1}{2}}FN^{-1}CN^{-1}F^TW^{\frac{1}{2}} & (29)\\
&= nobs - \mathrm{Tr}N^{-1}N_0 - \mathrm{Tr}N^{-1}CN^{-1}N_0 & (30)\\
&= nobs - nparam + \mathrm{Tr}N^{-1}CN^{-1}C & (31)
\end{aligned}$$

where extensive use has been made of the cyclic properties of the trace, i.e. TrAB =TrBA. In the absence of constraints we get the well known expectation for χ^2. The last term gives the contribution of the constraints. As expected, the constraints increase χ^2.

The equation that is used in solve is obtained from equation(25) by ignoring the last term:

$$W^{\frac{1}{2}}C_\Delta W^{\frac{1}{2}} = I_{nobs} - W^{\frac{1}{2}}FN^{-1}F^TW^{\frac{1}{2}} \tag{32}$$

taking the trace of this, we find:

$$\begin{aligned}
\chi^2_{solve} &= \mathrm{Tr}I_{nobs} - \mathrm{Tr}N^{-1}N_0 & (33)\\
&= nobs - \mathrm{Tr}N^{-1}(N-C) & (34)\\
&= nobs - nparam + \mathrm{Tr}N^{-1}C & (35)
\end{aligned}$$

The last term is fairly easy to calculate in solve. The difference between the correct answer, and the answer that solve uses is:

$$\begin{aligned}
\chi^2_{solve} - \chi^2 &= \mathrm{Tr}N^{-1}CN^{-1}N_0 & (36)\\
&= \mathrm{Tr}N_0^{\frac{1}{2}}N^{-1}CN^{-1}N_0^{\frac{1}{2}} & (37)\\
&\geq 0 & (38)
\end{aligned}$$

Hence the degrees of freedom calculated by solve are always too large. Since what is actually reported is χ^2 per degree of freedom, solve underestimates χ^2. The amount that it is underestimated is roughly 1-10%, with more typically values being in the lower end of this.

In the solve program *reway* I use equation (25) to calculate the expected value of the post fit residuals for various subsets. As presently implemented, I ignore the last term on the right hand side of this. The reason is that it is computationally expensive to calculate. Including this term slows down *cres* by about a factor of 5!